Towards the Rapid Development of a Natural Language Understanding Module

Catarina Moreira
catarina.p.moreira@ist.utl.pt

Ana Cristina Mendes
ana.mendes@l2f.inesc-id.pt

Luísa Coheur
luisa.coheur@l2f.inesc-id.pt

Bruno Martins
bruno.g.martins@ist.utl.pt

Instituto Superior Técnico, INESC-ID
Av. Professor Cavaco Silva, 2744-016 Porto Salvo, Portugal

The original publication is available at: Intelligent Virtual Agents, Springer
http://link.springer.com/chapter/10.1007%2F978-3-642-23974-8_33

Abstract

When developing a conversational agent, there is often an urgent need to have a prototype available in order to test the application with real users. A Wizard of Oz is a possibility, but sometimes the agent should be simply deployed in the environment where it will be used. Here, the agent should be able to capture as many interactions as possible and to understand how people react to failure. In this paper, we focus on the rapid development of a natural language understanding module by non experts. Our approach follows the learning paradigm and sees the process of understanding natural language as a classification problem. We test our module with a conversational agent that answers questions in the art domain. Moreover, we show how our approach can be used by a natural language interface to a cinema database.

1 Introduction

In order to have a clear notion of how people interact with a conversational agent, ideally the agent should be deployed at its final location, so that it can be used by people sharing the characteristics of the final users. This scenario allows the developers of the agent to collect corpora of real interactions. Although the Wizard of Oz technique [7] can also provide these corpora, sometimes it is not a solution if one needs to test the system with many different real users during a long period and/or it is not predictable when the users will be available.

The natural language understanding (NLU) module is one of the most important components in a conversational agent, responsible for interpreting the user requests. The

symbolic approach to NLU usually involves a certain level of natural language processing, which includes hand crafted grammars and requires a certain amount of expertise to develop them; by the same token, the statistical approach relies on a large quantity of labeled corpora, which is often not available.

In this paper we hypothesize that a very simple and yet effective NLU module can be built if we model the process of NLU as a classification problem, within the machine learning paradigm. Here, we follow the approach described in [5], although their focus is on frame-based dialogue systems. Our approach is language independent and does not impose any level of expertise to the developer: he/she simply has to provide the module with a set of possible interactions (the only constraint being the input format) and a dictionary (if needed). Given this input, each interaction is automatically associated with a virtual category and a classification model is learned. The model will map future interactions in the appropriate semantic representation, which can be a logical form, a frame, a sentence, etc. We test our approach in the development of a NLU module for EDGAR(Figure 1) a conversational agent operating in the art domain. Also, we show how the approach can be successfully used to create a NLU module for a natural language interface to a cinema database, JATEDIGO, responsible for mapping the user requests into logical forms that will afterwards be mapped into SQL queries[1].

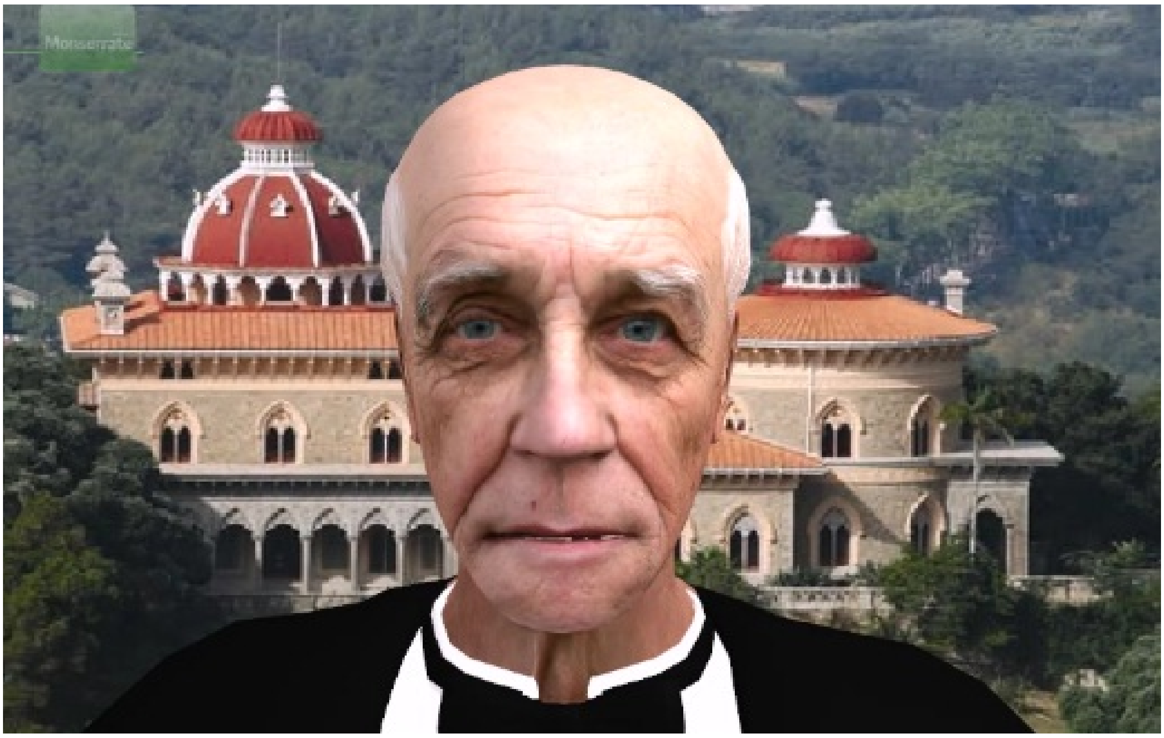

Figure 1: Agent Edgar

The paper is organized as follows: in Section 2 we present some related work and in Section 3 we describe our NLU module. Finally, in Section 4 we show our experiments and in Section 5 we conclude and present future work directions.

[1] All the code used in this work will be made available for research purposes at `http://qa.l2f.inesc-id.pt/`.

2 Related Work

NLU is the task of mapping natural language utterances into structures that the machine can deal with: the semantic representation of the utterances. The semantics of a utterance can be a logical form, a frame or a natural language sentence already understood by the machine. The techniques for NLU can be roughly split into two categories: symbolic and sub-symbolic. There are also hybrid techniques, that use characteristics of both categories.

Regarding symbolic NLU, it includes keyword detection, pattern matching and rule-based techniques. For instance, the virtual therapist ELIZA [11] is a classical example of a system based on pattern matching. Many early systems were based on a sophisticated syntax/semantics interface, where each syntactic rule is associated with a semantic rule and logical forms are generated in a bottom-up, compositional process. Variations of this approach are described in [2, 6]. Recently, many systems follow the symbolic approach, by using in-house rule-based NLU modules [4, 8]. However, some systems use the NLU modules of available dialogue frameworks, like the Let's Go system [10], which uses Olympus[2].

In what concerns sub-symbolic NLU, some systems receive text as input [5] and many are dealing with transcriptions from an Automatic Speech Recognizer [9]. In fact, considering speech understanding, the new trends considers NLU from a machine learning point of view. However, such systems usually need large quantities of labeled data and, in addition, training requires a previous matching of words into their semantic meanings.

3 The natural language understanding module

The NLU module receives as input a file with possible interactions (the training utterances file), from which several features are extracted. These features are in turn used as input to a classifier. In our implementation, we have used Support Vector Machines (SVM) as the classifier and the features are unigrams. However, in order to refine the results, other features can easily be included. Figure 2 describes the training phase of the NLU module.

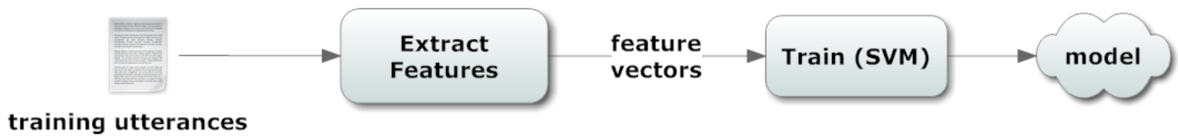

Figure 2: Training the NLU module.

Each interaction specified in the training utterances file is a pair, where the first element is a set of utterances that paraphrase each other and that will trigger the same response; the second element is a set of answers that represent possible responses to the previous utterances. That is, each utterance in one interaction represents different manners of expressing the same thing and each answer represents a possible answer to be returned by the system. The DTD of this file is the following:

[2] http://wiki.speech.cs.cmu.edu/olympus/index.php/Olympus.

```
<!ELEMENT corpus (interaction+)>
<!ELEMENT interaction (uterances, answers)>
<!ELEMENT utterances (u+)>
<!ELEMENT answers (a+)>
<!ELEMENT u (#PCDATA)>
<!ELEMENT a (#PCDATA)>
```

The NLU module also accepts as input a dictionary, containing elements to be replaced with labels that represent broader categories. Thus, and considering that TAG is the label that replaces a compound term $w_1 ... w_n$ during training, the dictionary is composed of entrances in the format:

TAG $w_1 ... w_n$ (for example: ACTOR Robert de Niro)

If the dictionary is used, Named Entity Recognition (NER) is performed to replace the terms that occur both in the training utterances file and user utterances. This process uses the LingPipe[3] implementation of the Aho-Corasick algorithm [1], that searches for matches against a dictionary in linear time in terms of the length of the text, independently of the size of the dictionary.

A unique identifier is then given to every paraphrase in each interaction – the interaction category – which will be the target of the training. For instance, since sentences *Há alguma data prevista para a conclusão das obras?* and *As obras vão acabar quando?* ask for the same information (*When will the conservation works finish?*), they are both labeled with the same category, generated during training: `agent_7`. The resulting file is afterwards used to train the classifier.

After the training phase, the NLU module receives as input a user utterance. If the NE flag is enabled, there is a pre-processing stage, where the NE recognizer tags the named entities in the user utterance before sending it to the classifier. Then the classifier chooses a category for the utterance. Since each category is associated with a specific interaction (and with its respective answers), one answer is randomly chosen and returned to the user. These answers must be provided in a file with the format `category answer`. Notice that more than one answer can be specified. Figure 3 describes the general pipeline of the NLU module.

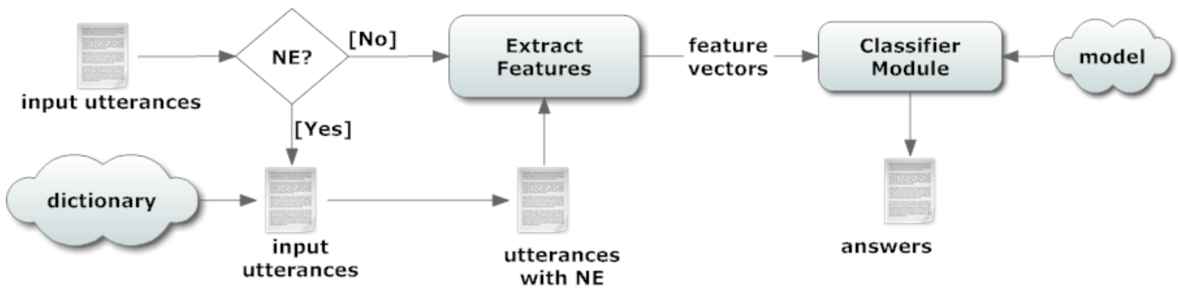

Figure 3: Pipeline of the NLU module.

[3]`http://alias-i.com/lingpipe/`.

4 Experiments

This section presents the validation methodology and the obtained results.

4.1 Experimental setup

In order to test our approach to the rapid development of a NLU module, we first collected a corpus that contains interactions in the art domain: the Art corpus. It was built to train Edgar, a conversational agent whose task is to engage in inquiry-oriented conversations with users, teaching about the Monserrate Palace. Edgar answers questions on its domain of knowledge, although it also responds to questions about himself. The Art corpus has 283 utterances with 1471 words, from which 279 are unique. The utterances represent 52 different interactions (thus, having each interaction an average of 5.4 paraphrases).

For our experiments in the cinema domain, we have used the Cinema corpus, containing 229 questions mapped into 28 different logical forms, each one representing different SQL queries. A dictionary was also build containing actor names and movie titles.

4.2 Results

The focus of the first experiment was to chose a correct answer to a given utterance. This scenario implies the correct association of the utterance to the set of its paraphrases. For instance, considering the previous example sentence *As obras vão acabar quando?*, it should be associated to the category `agent_7` (the category of its paraphrases).

The focus of the second experiment was to map a question into an intermediate representation language (a logical form) [3]. For instance, sentence *Que actriz contracena com Viggo Mortensen no Senhor dos Anéis?* (*Which actress plays with Viggo Mortensen in The Lord of the Rings?*) should be mapped into the form `WHO_ACTS_WITH_IN(Viggo Mortensen, The Lord of the Rings)`.

Both corpora where randomly split in two parts (70%/30%), being 70% used for training and 30% for testing. This process was repeated 5 times. Results are shown in Table 1.

Corpus	fold 1	fold 2	fold 3	fold 4	fold 5	average
Art	0.78	0.74	0.86	0.87	0.92	0.83
Cinema	0.87	0.90	0.79	0.77	0.82	0.83

Table 1: Accuracy results

4.3 Discussion

From the analysis of Table 1, we conclude that a simple technique can lead to very interesting results. Specially if we compare the accuracy obtained for the Cinema corpus with previous results of 75%, which were achieved with recourse to a linguistically rich framework that required several months of skilled labour to build. Indeed, the previous

implementation of JaTeDigo was based on a natural language processing chain, responsible for a morpho-syntactic analysis, named entity recognition and rule-based semantic interpretation.

Another conclusion is that one can easily develop an NLU module. In less than one hour we can have the set of interactions needed for training and, from there, the creation of the NLU module for that domain is straightforward. Moreover, new information can be easily added, allowing to retrain the model.

Nevertheless, we are aware of the debilities of our approach. The NLU module is highly dependent of the words used during training and the detection of paraphrases is only successful for utterances that share many words. In addition, as we are just using unigrams as features, no word is being detached within the input utterances, resulting in some errors. For instance, in the second experiment, the sentence *Qual o elenco do filme MOVIE?* (*Who is part of MOVIE's cast?*) was wrongly mapped into `QT_WHO_MAIN_ACT(MOVIE)`, although very similar sentences existed in the training. A solution for this problem is to add extra weight to some words, something that could be easily added as a feature if these words were identified in a list. Moreover, adding synonyms to the training utterances file could also help.

Another limitation is that the actual model does not comprise any history of the interactions. Also, we should carefully analyze the behavior of the system with the growing of the number of interactions (or logical forms), as the classification process becomes more complex.

5 Conclusions and Future Work

We have presented an approach for the rapid development of a NLU module based on a set of possible interactions. This approach treats the natural language understanding problem as a classification process, where utterances that are paraphrases of each other are given the same category. It receives as input two files, the only constraint being to write them in a given xml format, making it very simple to use, even by non-experts. Moreover, it obtains very promising results. As future work, and although moving from the language independence, we would like to experiment additional features and we would also like to try to automatically enrich the dictionary and the training files with relations extracted from WordNet.

Acknowledgments

This work was supported by FCT (INESC-ID multiannual funding) through the PIDDAC Program funds, and also through the project FALACOMIGO (Projecto em co-promoção, QREN n 13449). Ana Cristina Mendes is supported by a PhD fellowship from Fundação para a Ciência e a Tecnologia (SFRH/BD/43487/2008).

References

[1] Alfred V. Aho and Margaret J. Corasick. Efficient string matching: an aid to bibliographic search. *Communications of the ACM*, 18:333–340, June 1975.

[2] James Allen. *Natural language understanding (2nd ed.)*. Benjamin-Cummings Publishing Co., Inc., 1995.

[3] I. Androutsopoulos, G.D. Ritchie, and P. Thanisch. Natural language interfaces to databases–an introduction. *Journal of Language Engineering*, 1(1):29–81, 1995.

[4] Niels Ole Bernsen and Laila Dybkjær. Domain-Oriented Conversation with H.C. Andersen. In *Proc. of the Workshop on Affective Dialogue Systems, Kloster Irsee*, pages 142–153. Springer, 2004.

[5] Rahul Bhagat, A. Leuski, and Eduard Hovy. Shallow semantic parsing despite little training data. In *Proc. ACL/SIGPARSE 9th Int. Workshop on Parsing Technologies*, 2005.

[6] Daniel Jurafsky and James H. Martin. *Speech and Language Processing (2nd Edition)*. Prentice-Hall, Inc., 2006.

[7] J. F. Kelley. An iterative design methodology for user-friendly natural language office information applications. In *ACM Transactions on Office Information Systems*, 1984.

[8] Stefan Kopp, Lars Gesellensetter, Nicole C. Krämer, and Ipke Wachsmuth. *A conversational agent as museum guide: design and evaluation of a real-world application*, pages 329–343. Springer-Verlag, London, UK, 2005.

[9] Lucía Ortega, Isabel Galiano, Lluís F. Hurtado, Emilio Sanchis, and Encarna Segarra. A statistical segment-based approach for spoken language understanding. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association*, pages 1836–1839, 2010.

[10] Antoine Raux, Dan Bohus, Brian Langner, Alan W Black, and Maxine Eskenazi. Doing research on a deployed spoken dialogue system: One year of let's go! experience. In *Proceedings of the 7th Annual Conference of the International Speech Communication Association*, pages 65–68, 2006.

[11] Joseph Weizenbaum. Eliza - a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1):36–45, 1966.